EPI289 Spring 2023
Homework 1

TOTAL [38 points]

This homework is due on Wednesday, February 1, 2023, at 9:45 AM Eastern Time.

For multiple choice questions below, select the one best choice from the alternatives presented. For questions that may have multiple correct answers, the question will ask you to select all that apply.

For numerical answer questions, provide the number only (no letters, symbols, or other text). Use exact numbers and don't round until the last step of your calculation. Please keep in mind the units and number of decimal places requested in the question, if applicable. For short answer questions, please type your answers in the space provided.

You will need to use the hmwk1 and NHEFS dataset to complete this homework.

PART 1 [7.5 points]

The table below shows data collected on smoking *A* (1: yes, 0: no) and death *Y* (1: yes, 0: no) in a study of 410 individuals. For all questions in Part 1, round your answers to the nearest hundredth (i.e. 2 decimal places).

	A = 1	A = 0
Y = 1	168	62
Y = 0	44	136

1. **[Fill in the blank, 0.5 points] Calculate the odds of death among smokers.**

   Odds of death among smokers (2 decimal places) = __3.82**___ over the study period**

168/44

2. **[Fill in the blank, 0.5 points] Calculate the odds of death among nonsmokers.**

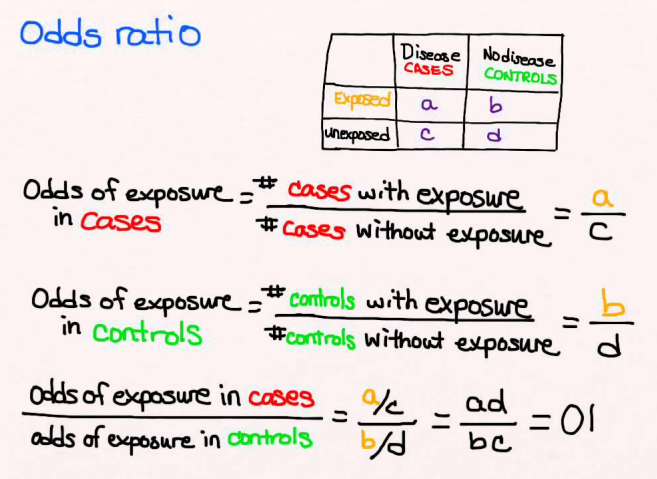
   Odds of death among nonsmokers (2 decimal places) = __0.46**__ over the study period**

62/136

3. **[Fill in the blank, 0.5 points] Calculate the odds ratio of death for smokers versus nonsmokers.**

   Odds ratio of death for smokers versus nonsmokers (2 decimal places) = _____ over the study period

(Pr(Y=1|A=1)/Pr[Y=1|A=0])/(Pr(Y=0|A=1)/Pr[Y=0|A=0]) = (168/62)/(44/136) = 8.38



4. **[Multiple choice, 1 point] Which of the following formulas did you use to calculate the odds ratio in Question 3?**
    a. Pr(Y=1|A=1)/Pr[Y=0|A=1]
    b. Pr(Y=1|A=0)/Pr[Y=0|A=0]
    c. Pr(Y=1|A=1)/Pr[Y=1|A=0]
    d. (Pr(Y=1|A=1)/Pr[Y=0|A=1])/(Pr(Y=1|A=0)/Pr[Y=0|A=0])
    e. (Pr(Y=1|A=0)/Pr[Y=0|A=0])/(Pr(Y=1|A=1)/Pr[Y=0|A=1])

d

For Questions 5 to 13, consider the saturated logistic model:

$$\textbf{logit Pr(Y=1|A=a)} = \alpha_0 + \alpha_1 A$$

Unless otherwise stated, do not use R or any other statistical software.

5. **[Fill in the blank, 0.5 points] What is the value of α_0?**

    _____ over the study period (2 decimal places)

logit Pr(Y=1|A=0) = α_0 = ln(62/410)=-1.89
logit Pr(Y=1|A=1) = α_0+ α_1

6. **[Multiple choice, 0.5 points] What is the interpretation of α_0?**
    a. Risk of death among smokers
    b. Odds of death among smokers
    c. Log odds of death among smokers
    d. Risk of death among nonsmokers
    e. Odds of death among nonsmokers
    f. Log odds of death among nonsmokers
    g. Risk ratio of death comparing smokers to nonsmokers
    h. Odds ratio of death comparing smokers to nonsmokers
    i. Log odds ratio of death comparing smokers to nonsmokers

d

7. **[Fill in the blank, 0.5 points] What is the value of α_1?**

    _____ over the study period (2 decimal places)

logit Pr(Y=1|A=1) = α_0+ α_1 **= ln(168/410)**
α_1 **= ln(168/410)-** ln(62/410)=ln(168*/62)=1.00= 0.99682959435*

8. **[Multiple choice, 0.5 points] What is the interpretation of α_1?**
    a. Risk of death among smokers

b. Odds of death among smokers
c. Log odds of death among smokers
d. Risk of death among nonsmokers
e. Odds of death among nonsmokers
f. Log odds of death among nonsmokers
g. Risk ratio of death comparing smokers to nonsmokers
h. Odds ratio of death comparing smokers to nonsmokers
i. Log odds ratio of death comparing smokers to nonsmokers

α_1 = logit Pr(Y=1|A=1)- logit Pr(Y=1|A=0) = logit$\frac{\Pr(Y=1|A=1)}{\Pr(Y=1|A=0)}$

i

9. [Fill in the blank, 0.5 points] What is the value of exp(α_0+α_1)?

__0.41__ over the study period (2 decimal places)

logit Pr(Y=1|A=1) = α_0+ α_1 = exp(ln(168/410)) = 168/41*0*

10. [Multiple choice, 0.5 points] What is the interpretation of exp(α_0+α_1)?

a. Risk of death among smokers
b. Odds of death among smokers
c. Log odds of death among smokers
d. Risk of death among nonsmokers
e. Odds of death among nonsmokers
f. Log odds of death among nonsmokers
g. Risk ratio of death comparing smokers to nonsmokers
h. Odds ratio of death comparing smokers to nonsmokers
i. Log odds ratio of death comparing smokers to nonsmokers

a

11. [Fill in the blank, 0.5 points] What is the value of exp(α_0+α_1)/exp(α_0) ?

__2.71__ over the study period (2 decimal places)

$\frac{\Pr(Y=1|A=1)}{\Pr(Y=1|A=0)}$ =(168/410)/(62/410)= 2.70967741935

12. [Multiple choice, 0.5 points] What is the interpretation of exp(α_0+α_1)/exp(α_0)?

a. Risk of death among smokers
b. Odds of death among smokers
c. Log odds of death among smokers
d. Risk of death among nonsmokers
e. Odds of death among nonsmokers
f. Log odds of death among nonsmokers
g. Risk ratio of death comparing smokers to nonsmokers

h. Odds ratio of death comparing smokers to nonsmokers
i. Log odds ratio of death comparing smokers to nonsmokers

H

13. [Fill in multiple blanks, 1 point] Use R to fit this logistic model to the dataset hmwk1.csv. What is the coefficient for SMK and its standard error?

Coefficient (2 decimal places): _____
Standard error (2 decimal places): _____

14. [Essay question, ungraded] Provide your R code for Part 1.

PART 2 [9.5 points]

The tables below show data collected in the same study of 410 individuals among drinkers (C=1) and nondrinkers (C=0). For all questions in Part 2, round your answers to the nearest hundredth (i.e. 2 decimal places).

C=1

	A = 1	A = 0
Y = 1	123	40
Y = 0	16	76

C=0

	A = 1	A = 0
Y = 1	45	22
Y = 0	28	60

15. **[Fill in multiple blanks, 0.5 points] Calculate the odds ratio of death for smokers versus nonsmokers among drinkers.**

(Pr(Y=1|A=1, C=1)/Pr[Y=1|A=0, C=1])/(Pr(Y=0|A=1, C=1)/Pr[Y=0|A=0, C=1]) = (123/40)/(44/136) = 8.38

OR (2 decimal places) = _____ over the study period

16. **[Fill in the blank, 0.5 points] Calculate the odds ratio of death for smokers versus nonsmokers among nondrinkers.**

OR (2 decimal places) = _____ over the study period

17. **[Multiple choice, 1 point] Which of the following formulas did you use to calculate the odds ratio of death for smokers compared to nonsmokers among drinkers?**
    a. Pr[Y=1|A=1,C=1]/Pr[Y=0|A=1,C=1]
    b. Pr[Y=1|A=1,C=1]/Pr[Y=1|A=1,C=0]
    c. Pr[Y=1|A=1,C=1]/Pr[Y=1|A=0,C=1]
    d. Pr[Y=1|A=1,C=1]/Pr[Y=0|A=1,C=1]/(Pr[Y=1|A=0,C=1]/Pr[Y=0|A=0,C=1])
    e. Pr[Y=1|A=1,C=1]/Pr[Y=1|A=0,C=1]/(Pr[Y=1|A=1,C=1]/Pr[Y=1|A=0,C=1])

For Questions 18 to 30, consider the saturated logistic model:

$$\text{logit } Pr(Y=1|A=a, C=c) = \alpha_0 + \alpha_1 A + \alpha_2 C + \alpha_3 AC$$

where AC is a product term between A and C. Unless otherwise stated, do not use R or any other statistical software.

18. [Fill in the blank, 0.5 points] What is the value of α_1?

_____ over the study period (2 decimal places)

19. [Multiple choice, 0.5 points] What is the interpretation of α_1?

a. Odds ratio of death comparing smokers to nonsmokers among drinkers
b. Log odds ratio of death comparing smokers to nonsmokers among drinkers
c. Odds ratio of death comparing smokers to nonsmokers among nondrinkers
d. Log odds ratio of death comparing smokers to nonsmokers among nondrinkers
e. Odds ratio of death comparing drinkers to nondrinkers among smokers
f. Log odds ratio of death comparing drinkers to nondrinkers among smokers
g. Odds ratio of death comparing drinkers to nondrinkers among nonsmokers
h. Log odds ratio of death comparing drinkers to nondrinkers among nonsmokers

20. [Fill in the blank, 0.5 points] What is the value of α_2?

_____ over the study period (2 decimal places)

21. [Multiple choice, 0.5 points] What is the interpretation of α_2?

a. Odds ratio of death comparing smokers to nonsmokers among drinkers
b. Log odds ratio of death comparing smokers to nonsmokers among drinkers
c. Odds ratio of death comparing smokers to nonsmokers among nondrinkers
d. Log odds ratio of death comparing smokers to nonsmokers among nondrinkers
e. Odds ratio of death comparing drinkers to nondrinkers among smokers
f. Log odds ratio of death comparing drinkers to nondrinkers among smokers
g. Odds ratio of death comparing drinkers to nondrinkers among nonsmokers
h. Log odds ratio of death comparing drinkers to nondrinkers among nonsmokers

22. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_2)$?

_____ over the study period (2 decimal places)

23. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_2)$?

a. Odds ratio of death comparing smokers to nonsmokers among drinkers

b. Log odds ratio of death comparing smokers to nonsmokers among drinkers
c. Odds ratio of death comparing smokers to nonsmokers among nondrinkers
d. Log odds ratio of death comparing smokers to nonsmokers among nondrinkers
e. Odds ratio of death comparing drinkers to nondrinkers among smokers
f. Log odds ratio of death comparing drinkers to nondrinkers among smokers
g. Odds ratio of death comparing drinkers to nondrinkers among nonsmokers
h. Log odds ratio of death comparing drinkers to nondrinkers among nonsmokers

24. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_2+\alpha_3)$?

_____ over the study period (2 decimal places)

25. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_2+\alpha_3)$?

a. Odds ratio of death comparing smokers to nonsmokers among drinkers
b. Log odds ratio of death comparing smokers to nonsmokers among drinkers
c. Odds ratio of death comparing smokers to nonsmokers among nondrinkers
d. Log odds ratio of death comparing smokers to nonsmokers among nondrinkers
e. Odds ratio of death comparing drinkers to nondrinkers among smokers
f. Log odds ratio of death comparing drinkers to nondrinkers among smokers
g. Odds ratio of death comparing drinkers to nondrinkers among nonsmokers
h. Log odds ratio of death comparing drinkers to nondrinkers among nonsmokers

26. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_1+\alpha_3)$?

_____ over the study period (2 decimal places)

27. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_1+\alpha_3)$?

a. Odds ratio of death comparing smokers to nonsmokers among drinkers
b. Log odds ratio of death comparing smokers to nonsmokers among drinkers
c. Odds ratio of death comparing smokers to nonsmokers among nondrinkers
d. Log odds ratio of death comparing smokers to nonsmokers among nondrinkers
e. Odds ratio of death comparing drinkers to nondrinkers among smokers
f. Log odds ratio of death comparing drinkers to nondrinkers among smokers
g. Odds ratio of death comparing drinkers to nondrinkers among nonsmokers
h. Log odds ratio of death comparing drinkers to nondrinkers among nonsmokers

28. [Fill in the blank, 0.5 points] What is the value of α_3?

_____ over the study period (2 decimal places)

29. **[Essay question, 1 point] What is the interpretation of α_3? Please limit your response to one sentence.**

alpha_3 is the difference between the two log odds ratio: the log odds ratio for death comparing smokers to nonsmokers among drinkers, the log odds ratio for death comparing smokers and nonsmokers among nondrinkers.

30. **[Fill in multiple blanks, 1 point] Use R to fit this logistic model to the dataset hmwk1.csv. What is the coefficient for SMK and its standard error?**

    Coefficient (2 decimal places): _____
    Standard error (2 decimal places): _____

31. **[Essay question, ungraded] Provide your R code for Part 2.**

PART 3 [9 points]

The tables below show the same data in overweight (B=1) and non-overweight (B=0) individuals. For any interpretation questions below, interpret the expressions as written with respect to these variables. For all questions in Part 3, round your answers to the nearest hundredth (i.e. 2 decimal places).

B=1

	A = 1	A = 0
Y = 1	60	38
Y = 0	10	62

B=0

	A = 1	A = 0
Y = 1	108	24
Y = 0	34	74

32. **[Fill in the blank, 0.5 points] Calculate the odds ratio of death for smokers versus nonsmokers among overweight individuals.**

    OR (2 decimal places) = _____ over the study period

33. **[Fill in the blank, 0.5 points] Calculate the odds ratio of death for smokers versus nonsmokers among non-overweight individuals.**

    OR (2 decimal places) = _____ over the study period

For Questions 34 to 48, consider the logistic model:

$$\text{logit } \Pr(Y=1|A=a, B=b) = \alpha_0 + \alpha_1 A + \alpha_2 B$$

Unless otherwise stated, do not use R or any other statistical software.

34. **[Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_2)$?**

    _____ over the study period (2 decimal places)

35. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_2)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight
e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals
h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

36. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_0)$?

_____ over the study period (2 decimal places)

37. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_0)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight
e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals
h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

38. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_0+\alpha_1)$?

_____ over the study period (2 decimal places)

39. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_0+\alpha_1)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight
e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals
h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

40. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_0+\alpha_1+\alpha_2)$?

_____ over the study period (2 decimal places)

41. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_0+\alpha_1+\alpha_2)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight
e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals

h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

42. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_0+\alpha_2)$?

_____ over the study period (2 decimal places)

43. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_0+\alpha_2)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight
e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals
h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

44. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_0+\alpha_1)/\exp(\alpha_0)$?

_____ over the study period (2 decimal places)

45. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_0+\alpha_1)/\exp(\alpha_0)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight

e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals
h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

46. [Fill in the blank, 0.5 points] What is the value of $\exp(\alpha_0+\alpha_1+\alpha_2)/\exp(\alpha_0+\alpha_2)$?

_____ over the study period (2 decimal places)

47. [Multiple choice, 0.5 points] What is the interpretation of $\exp(\alpha_0+\alpha_1+\alpha_2)/\exp(\alpha_0+\alpha_2)$?

a. Odds of death among individuals who are smokers and overweight
b. Odds of death among individuals who are smokers and non-overweight
c. Odds of death among individuals who are nonsmokers and overweight
d. Odds of death among individuals who are nonsmokers and non-overweight
e. Odds ratio of death comparing smokers to nonsmokers within levels of overweight status (i.e. among individuals who are overweight and among individuals who are non-overweight)
f. Odds ratio of death comparing smokers to nonsmokers among overweight individuals
g. Odds ratio of death comparing smokers to nonsmokers among non-overweight individuals
h. Odds ratio of death comparing overweight to non-overweight individuals within levels of smoking (i.e. among individuals who are smokers and among individuals who are nonsmokers)
i. Odds ratio of death comparing overweight to non-overweight individuals among smokers
j. Odds ratio of death comparing overweight to non-overweight individuals among nonsmokers

48. [Fill in multiple blanks, 1 point] Use R to fit this logistic model to the dataset hmwk1.csv. What is the coefficient for SMK and its standard error?

Coefficient (2 decimal places): _____
Standard error (2 decimal places): _____

49. [Essay question, ungraded] Provide your R code for Part 3.

PART 4 [12 points]

We are interested in the average causal effect of smoking cessation *A* (1: yes, 0: no) on the 10-year risk of death *D* (1: yes, 0: no), conditional on the following confounders *L*: sex (1: female, 0: male), race (1: nonwhite, 0; white), and age (continuous variable). For all questions in Part 4, round your answers to the nearest hundredth (i.e. 2 decimal places).

50. [Multiple choice, 1 point] Which of the following expressions represents this effect on the additive scale in counterfactual notation?

a. $Pr[D=1|A=1] - Pr[D=1|A=0]$
b. $Pr[D=1|A=1, L] - Pr[D=1|A=0, L]$
c. $Pr[D^{a=1}=1] - Pr[D^{a=0}=1]$
d. $Pr[D^{a=1}=1|L] - Pr[D^{a=1}=0|L]$
e. $Pr[D^{a=1}=1|L] - Pr[D^{a=0}=1|L]$
f. $Pr[D^{a=1}=1|A, L] - Pr[D^{a=0}=1|A, L]$

D

51. [Essay question, 2 point] Your uncle is a famous novelist with interest, but no formal training, in epidemiology. He does not understand what the counterfactual expression you just wrote means. Explain it to him in plain English. Please limit your response to one sentence.

A counterfactual expression describes a situation with a 'what if' situation, for instance, what would be the outcome of this disease had this patient be treated with medication A rather than B what s/he has been taking?

Estimate the average causal effect from question 50 in the NHEFS data, conditional on the following confounders L : sex (1: female, 0: male), race (1: nonwhite, 0; white), and age (continuous variable). Assume that the above 3 confounders are sufficient to guarantee conditional exchangeability. Model age with a linear and a quadratic term. Do not include any product terms in your model.

52. [Fill in multiple blanks, 1 point] Provide a point estimate and 95% confidence intervals for the causal odds ratio.

OR (2 decimal places): _____
Lower bound of 95% confidence interval (2 decimal places): _____
Upper bound of 95% confidence interval (2 decimal places): _____

53. [Multiple choice, 1 point] Did you use a parametric or a nonparametric approach?

a. Parametric

b. Nonparametric

54. [Multiple answers, 2 points] What are the modeling assumption(s) that you made, if any? Please note that some of these statements may represent equivalent modeling assumptions. (Select all that apply.)

a. The association between age and risk of death is curvilinear conditional on sex, race and quitting smoking.
b. The association between age and odds of death is curvilinear conditional on sex, race and quitting smoking.
c. The association between age and log odds of death is curvilinear conditional on sex, race and quitting smoking.
d. The contributions of sex, race, age, age2 and quitting smoking to the risk of death are additive.
e. The contributions of sex, race, age, age2 and quitting smoking to the odds of death are additive.
f. The contributions of sex, race, age, age2 and quitting smoking to the log odds of death are additive.
g. There are no interactions between age, sex, race and quitting smoking on the risk of death scale.
h. There are no interactions between age, sex, race and quitting smoking on the odds of death scale.
i. There are no interactions between age, sex, race and quitting smoking on the log odds of death scale.
j. There are no product terms between age, sex, race and quitting smoking on the risk of death scale.
k. There are no product terms between age, sex, race and quitting smoking on the odds of death scale.
l. There are no product terms between age, sex, race and quitting smoking on the log odds of death scale.
m. Conditional exchangeability, consistency, positivity and well-defined intervention.
n. No modeling assumption was made

55. [Essay question, 1 point] In the case of a binary outcome, why do you think people tend to use a logistic model to estimate odds ratio rather than a regular linear regression model to estimate the risk difference? Please limit your response to one sentence.

As logistic model transforms the output into the log-odds scale, which is easier to interpret and calculate the log odds ratio than the risk difference scale generated by the linear model.

56. [Multiple answers, 1 point] Your uncle pointed out that you may need to also adjust for years of smoking (smokeyrs, continuous). Which of the following statements provides a valid justification for considering years of smoking as an additional confounder? (Select all that apply.)

a. Years of smoking is associated with smoking cessation and associated with 10-year risk of death.
b. Years of smoking is a common cause of smoking cessation and 10-year risk of death.
c. Years of smoking can be used to block a backdoor path between smoking cessation and 10-year risk of death on a causal diagram.
d. Conditioning on years of smoking results in a change in the estimate for the association between smoking cessation and 10-year risk of death.

57. [Fill in multiple blanks, 1 point] Provide a point estimate and 95% confidence intervals for the causal odds ratio after including years of smoking (smokeyrs, continuous), sex (1: female, 0: male), race (1: nonwhite, 0; white), and age (continuous) in your model. Model age and smokeyrs with a linear and a quadratic term. Do not include any product terms in your model.

OR (2 decimal places): _____
Lower bound of 95% confidence interval (2 decimal places): _____
Upper bound of 95% confidence interval (2 decimal places): _____

58. [Multiple choice, 1 point] To consistently estimate the average causal effect in the population using outcome regression on the confounders, are you, in general, forced to make any assumption(s) about effect modification?

a. Yes
b. No

59. [Essay question, 1 point] If you selected "No" in question 58, explain why you did not need to make any assumptions. If you selected "Yes" in question 58, explain the assumption(s) in a way that your uncle can understand. Please limit your response to one sentence.

Effect modification can be present if the average causal effect varies within levels of confounders. Here, we make an assumption while modeling that there is no effect of confounder modification such that the estimated causal odds ratio can be constant across the level of confounders.

60. [Essay question, ungraded] Provide your R code for Part 4.

PART 5 (OPTIONAL - ungraded)

Make sure you know how to carry out these tasks in R. Use the NHEFS data to explore the relation between age and weight gain.

61. **Generate a temporary dataset that excludes observations with missing values for weight gain and create the following categories of age (25-40, 41-60, and >60).**

62. **Use *cut* in R to generate categorical variables with 5, 10, 20 and 49 categories of age.**

63. **Use *plot* in R to graphically explore the relation between age and weight gain.**

64. **Fit a linear regression model of the form $wt82\_71 = \alpha 0 + \alpha 1 Age$ with age as a continuous variable and plot the values predicted by the model against the observed values.**

65. **Similarly, fit a linear regression model of the form $wt82\_71 = \alpha 0 + \alpha 1 Age + \alpha 2 Age^2$ and plot the values predicted by the model against the observed values.**

66. **Fit a linear regression model including age in 5, 10, 20 and 49 categories and plot the values predicted by the model against the observed values.**